Algorithm of Successives Restrictions

Linda Smail
Marne-la-Vallée University
France
smail@math.univ-mlv.fr

Abstract

Given a bayesian network relative to a set I of discrete random variables, we are interested in computing the probability distribution $P(A/B)$, where A and B are two disjoined subsets of I. The general idea of the algorithm of successives restrictions is to manage the successions of summations on all random variables out of the target in order to keep on the target a structure less constraining than the Bayesian network, but which allows saving in memory ; that is the structure of Bayesian Network of Level Two.

1 Introduction

Given a bayesian network relative to a set $(X_i)_{i\in I}$ of discrete random variables, we are interested in computation of the probability distribution of a subset A of random variables conditionally to another subset B, where A and B are two disjoined subsets of I. According to Bayes theorem : $P(x_A|x_B) = \frac{P(x_A, x_B)}{\sum_{x_A} P(x_A, x_B)}$ thus to compute this conditional probability we need to compute the probability distribution of $X_{A\cup B}$, which we called "target". The general idea is to manage the successions of summations on all random variables X_i, for $i \in (I - (A \cup B)) = \overline{A \cup B}$.

Our aim is to compute $P(x_A|x_B)$, for one target A, not as in many studies which developed some effective algorithms to compute $P(x_i|x_B)$ for each $i \in I$, this makes the difference between our algorithm and the clustering-based algorithms. In this algorithm we do not state on the target any particular assumption in connection with the structure of the bayesian network. The subset A can have more than one variable, and not specially variables which belong to the same clique as in (Lauritzen, 1988) where the targets are cliques organized in a structure of a junction tree. In other words, in (Lauritzen, 1988), we can compute $P(x_A|x_B)$ if A is a subset of variables of the same clique.

2 Bayesian network of level two

We consider a probability P_I of a finite family $(X_i)_{i\in I}$, of random variables on a finite space Ω_I. Let $\mathcal{I}$ be a partition of I and let us consider a directed acyclic graph $\mathcal{G}$ on $\mathcal{I}$; we say that there is a link from $J^{'}$ to $J^{''}$ (where $J^{'}$ and $J^{''}$ are atoms of the partition $\mathcal{I}$) if $(J^{'}, J^{''}) \in \mathcal{G}$. If $J \in \mathcal{I}$, we note $p(J)$ the set of parents of J, that is the set of $J^{'}$ such that $(J^{'}, J) \in \mathcal{G}$.
Let $a(J)$ be the *initial subset* (ancestors) defined by J, in other words the set consisting in J itself and the J'' such as there is a path in $\mathcal{G}$ from J'' to J ; we can identify it with the union of all J'' such that $J'' \in a(J)$.
Definition 1 : The probability P_I is defined by the *Bayesian network of level two* (BN2), on I, $(\mathcal{I}, G, (P_{J/p(J)})_{J\in\mathcal{I}})$, if for each $J \in \mathcal{I}$, we have the conditional probability $P_{J/p(J)}$, in other words the probability of X_J conditioned by $X_{p(J)}$ (which, if $p(J) = \emptyset$, is the marginal probability P_J), such as :

$$P_I(x_I) = \prod_{J\in\mathcal{I}} P_{J/p(J)}(x_J/x_{p(J)}).$$

An usual Bayesian network (said *of level one* : BN1) is a particular case of BN2, with the partition of I into singletons.

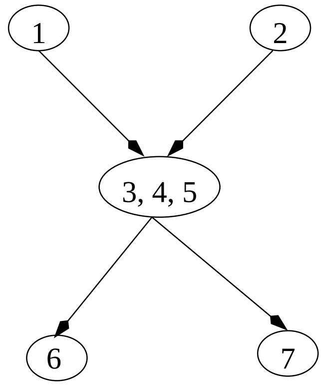

Figure 1: *Example of a Bayesian network of level two*

Definition 2 : Let $cd(J)$ be the set (possibly empty) of the *close descendants* of J, in other words the children of J and, if there are any out of the children themselves, the vertices located on a path between J and one of his children.

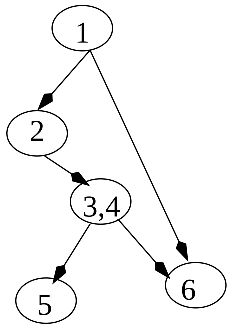

Figure 2: *Bayesian network of level two*

$cd(1) = \{2, \{3, 4\}, 6\}$, which we can write also : $cd(1) = \{2, 3, 4, 6\}$.

3 Algorithme of successives restrictions

Let $(I, G, (P_{i/p(i)})_{i \in I})$ be a bayesian network. Given a subset A of I, known as "target", we consider the problem of computing the probability distribution of $(X_i)_{i \in A} = X_A$. To solve this problem, we suggest an algorithm said *algorithm of successives restrictions*. The general idea is to manage the summations that we have to do relatively to the random variables X_ℓ for $\ell \in \overline{A}$ ($\overline{A} = I - A$).

3.1 Principele of the algorithme

Given a bayesian network $(I, G, (P_{i|p(i)})_{i \in I})$ and a subset A of I, the general idea of the algorithm of successive restrictions is to build a sequence of subsets $(I_0, \dots, I_\ell)$ with $\ell = \text{Card } (I - A)$, and for each $0 \le s \le \ell$ a structure of bayesian network of level 2 on I_s, noted $R_s = (\mathcal{I}_s, G_s, (P_{J|p(J)})_{J \in \mathcal{I}_s}$, which defines $P_{\mathcal{I}_s}$, probability ditribution of $X_{I_s} = (X_i)_{i \in I_s}$, such that :

1. $\mathcal{I}_0$ is the initial network (so $I_0 = I$).

2. Each element of I_s which contains an element of $\overline{A}$ is a singleton.

3. Once the algorithm is performed, $I_\ell = A$ and the probability distribution of $(X_i)_{i\in A}$ can be computed simply by product of the conditional probabilities in the bayesian network of level two obtained on I_ℓ .

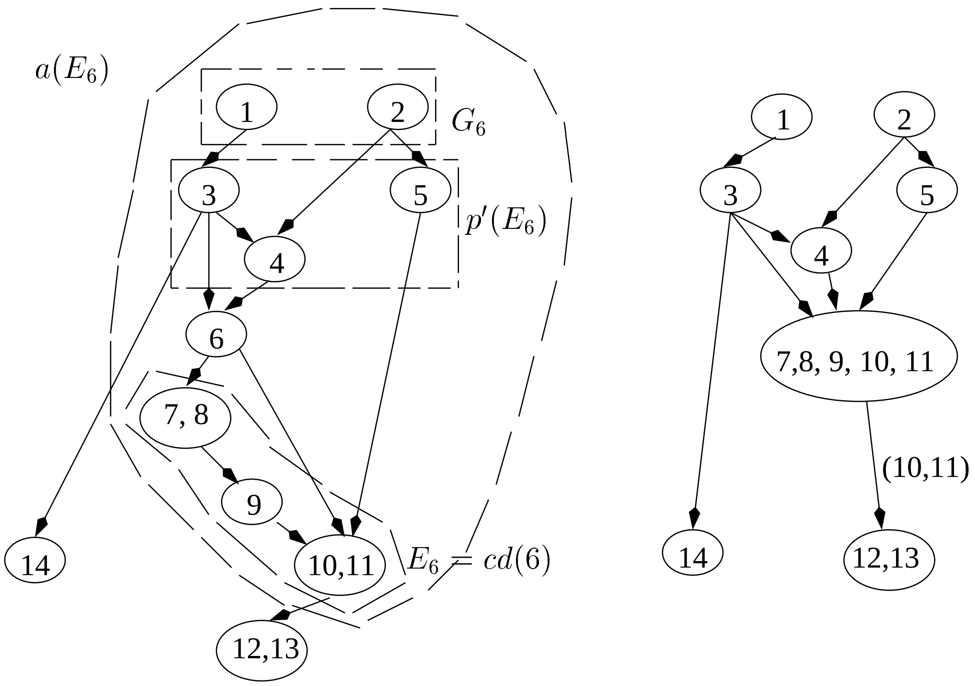

Figure 3: *(a) : Example of BN2. (b) : BN2 resulting after summation over* 6 *:* $E_6 = \{7, 8, 9, 10, 11\}$.

Now let's specify the general step of this algorithm : we start with a BN2, on $L \subset I$, $R = (\mathcal{I}, G, (P_{J/p(J)})_{J\in\mathcal{I}})$, defining P_L, where $L = \cup_{J\in\mathcal{I}} J$; we will obtain at the exit a BN2, on $L' = L - \{i\}$, $R' = (\mathcal{I}', G', (P_{J/p(J)})_{J\in\mathcal{I}'})$, defining $P_{L'}$. We choose a variable i of $\overline{A} \cap L$; according to what precedes, the singleton i belongs to $\mathcal{I}$; this choice is made in such a way that i has no descendant in $\overline{A}$. Than the BN2 R' has for vertices the atoms of the partition $\mathcal{I}'$, which are :

1. the set of close descendants of i in R, which we denote E_i ($E_i = cd(i)$),

2. all the atoms of $\mathcal{I}$ other than i and those in E_i .

The graph G' results from the graph G in the following way :

1. we delete the links (for G) relative to i and to the elements of E_i,

2. we keep all other links of G,

3. we introduce to E_i a set of parents, $p'(E_i)$, which includes : all parents of i in G and all parents of vertices belonging to the close desendants of i, others than i or those in E_i itself.

4. we introduce as children to E_i all children of vertices of R belonging to E_i, others than those in E_i itself.

The probabilistic data associated to R' can be computed from those associated to R in the following way :

1. we conserve the probability, conditionally to his parents, for each vertex such that the passage from R to R' changes neither itself nor his parents (in other words, each vertex other than i, those in E_i and the children of those in E_i);

2. for each child J of E_i (in G), his probability, conditionally to his parents, is preserved by substitution of E_i to the set of the parents of J (in G) which belongs to E_i, and we conserve the information that only these variables intervene in p'_{J/E_i}.

3. we create the probability of E_i conditionally to $p'(E_i)$, which can be computed owing to the following formula

$$P_{E_i/p'(E_i)}(x_{E_i}/x_{p'(E_i)}) = \sum_{x_i \in \Omega_i} \left[\left(\prod_{J \in E_i} P_{J/p(J)}(x_J/x_{p(J)}) \right) P_{i/p(i)}(x_i/x_{p(i)}) \right].$$

Proof :

To simplify the notation, we will note $\sum_i$ for $\sum_{x_i \in \Omega_i}$ and, for each subset $B = \{b_1, \ldots, b_m\}$ of L, $\sum_B$ for $\sum_{x_{b_1} \in \Omega_{b_1}} \cdots \sum_{x_{b_m} \in \Omega_{b_m}}$; we will omit to write the variables (for example we write $P_{J/p(J)}$ for $P_{J/p(J)}(x_J/x_{p(J)})$). The various objects intervening in this proof may be visualized on the example given in figure 3 (a).

The computation of $P_{E_i/p'(E_i)}$ can be done by using only vertices in $a(E_i)$ the initial subset defined by E_i, which includes $p'(E_i)$ by construction ; we decompose $a(E_i)$ according to the partition $(E_i, \{i\}, p'(E_i), G_i)$, where $a\Big(p'(E_i)\Big) = p'(E_i) \cup G_i$. Then

$$P_{E_i \cup p'(E_i)} = \sum_i \sum_{G_i} \left[\left(\prod_{J \in E_i} P_{J/p(J)} \right) P_{i/p(i)} \left(\prod_{k \in p'(E_i)} P_{k/p(k)} \right) \left(\prod_{h \in G_i} P_{h/p(h)} \right) \right].$$

We notice that the index i is present only in $\left(\prod_{J \in E_i} P_{J/p(J)} \right) P_{i/p(i)}$ whereas all the ℓ in G_i may be present only in $\left(\prod_{k \in p'(E_i)} P_{k/p(k)} \right) \left(\prod_{h \in G_i} P_{h/p(h)} \right)$ so

$$P_{E_i \cup p'(E_i)} = \left\{ \sum_i \left[\left(\prod_{J \in E_i} P_{J/p(J)} \right) P_{i/p(i)} \right] \right\} \left\{ \sum_{G_i} \left[\left(\prod_{k \in p'(E_i)} P_{k/p(k)} \right) \left(\prod_{h \in G_i} P_{h/p(h)} \right) \right] \right\}.$$

Since $P_{p'(E_i)} = \sum_{G_i} \left[\left(\prod_{k \in p'(E_i)} P_{k/p(k)} \right) \left(\prod_{h \in G_i} P_{h/p(h)} \right) \right]$, we get

$$P_{E_i/p'(E_i)} = \sum_i \left[\left(\prod_{J \in E_i} P_{J/p(J)} \right) P_{i/p(i)} \right].$$

References

Finn Jensen (1999). *An Introduction to Bayesian Networks.* UCL Press.

Pearl, J. (1986). Fusion propagation and structuring in belief networks. *Artificial Intelligence 29*(3), 241–288.

Steffen Lauritzen, David Spiegelhalter (1988). Local Computation with Probabilities on Graphical Structures and their Application to Expert Systems. Proceedings of the Royal Statistical Society, Series B 50 (2).